Constrained Cost-Coupled Stochastic Games with Independent State Processes*

Eitan Altman•, Konstantin Avrachenkov•, Nicolas Bonneau•,†,
Merouane Debbah†, Rachid El-Azouzi◇, Daniel Sadoc Menasche•,

•INRIA, Centre Sophia-Antipolis, 2004 Route des Lucioles, B.P.93,
06902 Sophia-Antipolis Cedex, France

†Mobile Communications Group, Institut Eurecom, 2229,
Route des Cretes, B.P. 193, 06904, Sophia Antipolis Cedex, France

◇LIA, Univesite d'Avignon, 339, chemin des Meinajaries,
Agroparc BP 1228, 84911 AVIGNON Cedex 9, FRANCE

arXiv:cs/0703099v1 [cs.IT] 21 Mar 2007

Abstract

We consider a non-cooperative constrained stochastic games with N players with the following special structure. With each player i there is an associated controlled Markov chain MDP_i. The transition probabilities of the ith Markov chain depend only on the state and actions of controller i. The information structure that we consider is such that each player knows the state of its own MDP and its own actions. It does not know the states of, and the actions taken by other players. Finally, each player wishes to minimize a time-average cost function, and has constraints over other time-avrage cost functions. Both the cost that is minimized as well as those defining the constraints depend on the state and actions of all players. We study in this paper the existence of a Nash equilirium. Examples in power control in wireless communications are given.

1 Introduction

Non-cooperative games deal with a situation of several decision makers (often called agents, users or players) where the cost of each one of the players may be a function of not only its own decision but also of decisions of other players. The choice of a decision by any player is done so as to minimize its own individual cost.

Non-cooperative games also allow to model sequential decision making by non-cooperating players. They allow to model situations in which the parameters defining the games vary in time. The game is then said to be a *dynamic game* and the parameters that may vary in time are the *states* of the game. At any given time (assumed to be discrete) each player takes a decision (also called an *action*) according to some strategy. The vector of actions chosen by players at a given time (called a *multi-action* may determine not only the cost for each player at that time; it can also determine the state evolution. Each player is interested in minimizing some functions of all the costs at different time instants. In particular, we shall consider here the expected time-average costs for the players.

We consider in this paper the class of stochastic decentralized games which we call "cost coupled constrained stochastic games" and are characterized by the following:

1. We associate to each player a Markov chain, whose transition probabilities depend only on the action of that player,

2. We assume that at any time, each player has information only on the current and past states of his own Markov chain as well as of his previous actions. It does not know the state and actions of other players.

3. Each player has constraints on its strategies (to be defined later). We consider the general situation in which the constraints for a player depend on the strategies used by other players.

*This work was supported by the Bionets European project

4. There are cost functions (one per player) that depend on the states and actions of *all* players, and each player wishes to minimize its own cost.

We see that players "interact" only through the last two points above.

It is well known that identifying equilibrium policies (even in absence of constraints) is hard. Unlike the situation in Markov Decision Processes (MDPs) in which stationary optimal strategies are known to exist (under suitable conditions), and unlike the situation in constrained MDPs (CMDPs) with a multichain structure, in which optimal Markov policies exist [13, 18], we know that equilibrium strategies in stochastic games need in general to depend on the whole history (see e.g. [19] for the special case of zero-sum games). This difficulty has motivated researchers to search for various possible structures of stochastic games in which saddle point policies exist among stationary or Markov strategies and are easier to compute [11]. In line with this approach, we shall identify conditions under which constrained equilibria exist for cost-coupled conostrained stochastic games.

Related work. Several papers have already dealt with constrained stochastic games. In [7], the authors have established the existence of a constnrained equilibrium in a context of centralized stochastic games, in which all players jointly control a single Markov chain and in which all players have full information on its state. Moreover, when taking decision at time t, each player has information on all actions previously taken by all players.

The special cost-coupled structure (see Definion 2.1) has been investigated in [12, 2] in *zero-sum games* where there is a single cost which one of the players wishes to minimize and which a second player wishes to maximize. A highly non-stationary saddle-point was obtained in [22] for a zero-sum constrained stochastic games with expected average costs.

Alghough the question of existence of an equilibrium in cost-coupled stochastic games has not been considered before, some specific applications of such games have been formulated. Indeed, these games have been used extensively by Huang, Malhamé and Caines in a series of publications [16, 17]. Although they have not established the existence of a Nash equilibrium, they have been able to obtain an ϵ-Nash equilibrium for the case of a large population of players. Models concerning uplink power control, similar to the one studied in [16], have been investigated in [3], in which the structure of constrained equilibrium is established. We note however that in the models considered in [3], the local Markovian states of each user are not controlled; the decisions of each user have an impact only the costs and not the transition probabilities.

2 The model and main result

We consider a game with N players, labeled $1, \ldots, N$. Define for each player i the tuple $\{\mathbf{X}_i, \mathbf{A}_i, \mathcal{P}_i, c_i, V_i, \beta_i\}$ where

- $\mathbf{X}_i$ is a finite **local** state space of the ith player. Generic notation for states will be x, y or x_i, y_i. We let $\mathbf{X} := \prod_{j=1}^{N} \mathbf{X}_j$ be the **global** state space, and we define $\mathbf{X}_{-i} := \prod_{j \neq i} \mathbf{X}_i$ be the **global** to be the set of all possible states of players other than i.

- $\mathbf{A}_i$ is a finite set of actions. We denote by $\mathbf{A}_i(x_i)$ the set of actions available for player i at state x. A generic notation for a vector of actions will be $\mathbf{a} = (a_1, ..., a_N)$ where a_i stands for the action chosen by player i.

- Define the local set of state-action pairs for player i as set $\mathcal{K}_i = \{(x_i, a_i) : x_i \in \mathbf{X}_i,\ a_i \in \mathbf{A}_i(x)\}$. Denote the set of all global state-action pairs by $\mathcal{K} = \prod_{i=j}^{N} \mathcal{K}_j$, and let $\mathcal{K}_{-i} = \prod_{j \neq}^{N} \mathcal{K}_j$ denote the set of state-action pairs of all players other than i.

- $\mathcal{P}^i$ are the transition probabilities for player i; thus $\mathcal{P}^i_{x_i a_i y_i}$ is the probability that the state of player i moves from x_i to y_i if she chooses action a_i.

- $c = \{c_i^j\}, i = 1, ..., N,\ j = 0, 1, ..., B_i$ is a set of immediate costs, where $c_i^j : \mathcal{K} \to \mathbb{R}$. Thus player i has a set of $B_i + 1$ immediate costs; c_i^0 will correspond to the cost function that is to be minimized by that player, and c_i^j, $j > 0$ will correspond to cost functions on which some constraints are imposed.

- $V = \{V_i^j\}, i = 1, ..., N,\ j = 1, ..., B_i$ are bounds defining the constraints (see (2) below).

- β_i is a probability distribution for the initial state of the Markov chain of player i. The intial states of the players are assumed to be independent.

Histories, Information and policies. Let $M_1(G)$ denote the set of probability measures over a set G. Define a history of player i at time (or of length) t to be a sequence of her previous states and actions, as well as her current local state: $h_i^t = (x_i^1, a_i^1, ..., x_i^{t-1}, a_i^{t-1}, x_i^t)$ where $(x_i^s, a_i^s) \in \mathcal{K}_i$ for all $s = 1, ..., t$. Let $\mathbf{H}_i^t$ be the set of all possible histories of length t for player i. A policy (also called a strategy) u_i for player i is a sequence $u_i = (u_i^1, u_i^2, ...)$ where $u_i^t : \mathbf{H}_i^t \to M_1(\mathbf{A}_i)$ is a function that assigns to any history of length t a probability measure over the set of actions of player i.

At time t, each player i chooses an action a_i, independently of the choice of actions of other players, with probability $u_i^t(a_i|h_i^t)$ if the history h_i^t was observed by player i. Denote $\mathbf{a} = (a_1, ..., a_N)$.

The class of all policies defined as above for player i is denoted by U^i. The collection $U = \prod_{i=1}^N U^i$ is called the class of multi-policies ($\prod$ stands for the product space).

Stationary policies. A stationary policy for player i is a function $u_i : \mathbf{X}_i \to M_1(\mathbf{A}_i)$ so that $u_i(\cdot|x_i) \in M_1(\mathbf{A}_i(x_i))$. We denote the class of stationary policies of player i by U_i^S. The set $U_S = \prod_{i=1}^N U_i^S$ is called the class of stationary multi-policies. Under any stationary multi-policy u (where the u^i are stationary for all the players), at time t, the controllers, independently of each other, choose actions $\mathbf{a} = (a_1, ..., a_N)$, where action a_i is chosen by player i with probability $u_i(a_i|x_i^t)$ if state x_i^t was observed by player i at time t.

For $u \in U$ we use the standard notation u_{-i} to denote the vector of policies $u_k, k \neq i$; moreover, for $v_i \in U_i$, we define $[u_{-i}|v_i]$ to be the multi-policy where, for $k \neq i$, player k uses u_k, while player i uses v_i. Define $U^{-i} := \cup_{u \in U}\{u_{-i}\}$.

A distribution β for the initial state (at time 1) and a multi-policy u together define a probability measure P_β^u which determines the distribution of the vector stochastic process $\{X^t,\ A^t\}$ of states and actions, where $X^t = \{X_i^t\}_{i=1,...,N}$ and $A^t = \{A_i^t\}_{i=1,...,N}$. The expectation that corresponds to an initial distribution β and a policy u is denoted by E_β^u.

Costs and constraints. For any multi-policy u and β, define the i,j-expected average cost is defined as

$$C^{i,j}(\beta, u) = \overline{\lim_{T\to\infty}} \frac{1}{T} \sum_{t=1}^{T} E_\beta^u c_i^j(X_t, A_t). \tag{1}$$

A multi-policy u is called i-feasible if it satisfies:

$$C^{i,j}(\beta, u) \leq V_i^j, \quad \text{for all } j = 1, ..., B_i. \tag{2}$$

It is called feasible if it is i-feasible for all the players $i = 1, ..., N$. Let U_V be the set of feasible policies.

Definition 2.1 *(i) A multi-policy $u \in U^v$ is called constrained Nash equilibrium if for each player $i = 1, ..., N$ and for any v_i such that $[u_{-i}|v_i]$ is i-feasible,*

$$C^{i,0}(\beta, u) \leq C^{i,0}(\beta, [u_{-i}|v_i]). \tag{3}$$

Thus, any deviation of any player i will either violate the constraints of the ith player, or if it does not, it will result in a cost $C^{i,0}$ for that player that is not lower than the one achieved by the feasible multi-policy u.
(ii) For any multi-policy u, u_i is called an optimal response for player i against u_{-i} if u is i-feasible, and if for any v^i such that $[u_{-i}|v_i]$ is i-feasible, (3) holds.
(iii) A multi-policy v is called an optimal response against u if for every $i = 1, ..., N$, v_i is an optimal response for player i against u_{-i}.

Assumptions. We introduce the following assumptions

- (Π_1) Ergodicity: For each player i and for any stationary policy u_i of that player, the state process of that player is an irreducible Markov chain with one ergodic class (and possibly some transient states).
- (Π_2) Strong Slater condition: There exists some real number $\eta > 0$ such that the following holds. Every player i has some policy v_i such that for any multi-strategy u_{-i} of the other players,

$$C^{i,j}(\beta, ([u_{-i}|v_i]) \leq V_i^j - \eta, \quad \text{for all } j = 1, ..., B_i. \tag{4}$$

- (Π_3) Information: The strategy chosen by any player does not depend on the realization of the cost.

The last assumption is frequently encountered in game theory and in applications, see e.g. [9, 21, 23]. The assumption is in fact directly implied by the definition of policies. If it were allowed to have policies depend on the realization of the cost, then a player could use the costs to estimate the state and actions of the other player.

We are now ready to introduce the main result.

Theorem 2.1 *Assume that Π_1 and Π_2 hold. Then there exists a stationary multi-policy u which is constrained-Nash equilibrium.*

Remark 2.1 *If assumption Π_2 does not hold, the upper semi-continuity which is needed for proving the existence of an equilibrium (see Proposition 3.1) need not hold. This is true even for the case of a single player, see [4].*

3 Proof of main result

We begin by describing the way an optimal stationary response for player i is computed for a given stationary multi-policy u. Fix a stationary policy u_i for player i. With some abuse of notation, we denote for any $x_i \in \mathbf{X}_i$ and any $y_i \in \mathbf{X}_i$,

$$\mathcal{P}^i_{x_i u_i y_i} = \sum_{a_i \in \mathbf{A}_i(x_i)} u_i(a_i|x_i) \mathcal{P}^i_{x_i a_i y_i}.$$

Denote the immediate costs induced by players other than i, when player i uses action a_i and the other players use a stationary multi policy u_{-i}, by

$$c_i^{j,u}(x_i, a_i) := \sum_{(\mathbf{x},\mathbf{a})_{-i} \in \mathcal{K}_{-i}} \left[\prod_{l \neq i} u_l(a_l|x_l) \pi_l^u(x_l) \right] c_i^j(\mathbf{x}, \mathbf{a}) \qquad \mathbf{a} = [\mathbf{a}_{-i}|a_i], \quad \mathbf{x} = [\mathbf{x}_{-i}|x_i],$$

Next we present a Linear Program (LP) for computing the set of all optimal responses for player i against a stationary policy u_{-i}.

LP(i, u) :
Find $\mathbf{z}^*_{\mathbf{i},\mathbf{u}} := \{z^*_{i,u}(y, a)\}_{y,a}$, where $(y, a) \in \mathcal{K}_i$, that minimizes

$$\mathcal{C}_u^{i,0}(z_i) := \sum_{(y,a) \in \mathcal{K}_i} c_i^{0,u}(y, a) z_{i,u}(y, a) \qquad \text{subject to:} \tag{5}$$

$$\sum_{(y,a) \in \mathcal{K}_i} z_{i,u}(y, a) \left[\delta_r(y) - \mathcal{P}^i_{yar} \right] = 0, \qquad \forall r \in \mathbf{X}_i, \tag{6}$$

$$\mathcal{C}_u^{i,j}(z_{i,u}) := \sum_{(y,a) \in \mathcal{K}_i} c_i^{j,u}(y, a) z_{i,u}(y, a) \leq V_i^j \qquad 1 \leq j \leq B_i \tag{7}$$

$$z_{i,u}(y, a) \geq 0, \quad \forall (y, a) \in \mathcal{K}_i \qquad \sum_{(y,a) \in \mathcal{K}_i} z_{i,u}(y, a) = 1 \tag{8}$$

Define $\Gamma(i,u)$ to be the set of optimal solutions of $\mathbf{LP}(i,u)$.

Given a set of nonnegative real numbers $z_i = \{z_i(y,a), (y,a) \in \mathcal{K}_i(y)\}$, define the point to set mapping $\gamma(i,z_i)$ as follows: If $\sum_a z_i(y,a) \neq 0$ then $\gamma_y^a(i,z_i) := \{z_i(y,a)[\sum_a z_i(y,a)]^{-1}\}$ is a singleton: for each y, we have that $\gamma_y(z_i) = \{\gamma_y^a(z_i) : a \in \mathbf{A}_i(y)\}$ is a point in $M_1(\mathbf{A}_i(y))$. Otherwise, $\gamma_y(i,z) := M_1(\mathbf{A}_i(y))$, i.e. the (convex and compact) set of all probability measures over $\mathbf{A}_i(y)$.

Define $g^i(z_i)$ to be the set of stationary policies for player i that choose, at state y_i, action a with probability in $\gamma_y^a(i,z_i)$.

For any stationary multi-policy v define the occupation measures

$$f(\beta,v) := \{f_i(v_i;y_i,a_i) : (y_i,a_i) \in \mathcal{K}_i,\ i = 1,...,N\}$$

as follows. Let

$$f_i(v_i;y_i,a_i) := \pi_i^{v_i}(y)v_i(a_i|y_i),$$

where $\pi_i^{v_i}$ is the steady state (invariant) probability of the Markov chain describing the state process of player i, when her policy is v_i. Note that a unique steady state probability exists by Assumption Π_1 and it does not depend on β. We thus often omit β from the notation.

Proposition 3.1 *Assume Π_1-Π_3. Fix any stationary multi-policy u.*
(i) If $z_{i,u}^*$ is an optimal solution for $\mathbf{LP}(i,u)$ then any element w in $g^i(z_{i,u}^*)$ is an optimal stationary response of i against the stationary policy u_{-i}. Moreover, the multi-policy $v = [u_{-i}|w]$ satisfies $f_i(v) = z_{i,u}^*$ (it does not depend on β).
(ii) Assume that w is an optimal stationary response of player i against the stationary policy u_{-i}, and let $v := [u_{-i}|w]$. Then $f_i(v)$ does not depend on β and is optimal for $\mathbf{LP}(i,u)$. i
(iii) The optimal sets $\Gamma(i,u)$, $i = 1,...,N$ are convex, compact, and upper semi-continuous in u_{-i}, where u is identified with points in $\prod_{i=1}^N \prod_{x_i \in \mathbf{X}_i} M_1(\mathbf{A}_i(x_i))$.
(iv) For each i, $g^i(z)$ is upper semi-continuous in z over the set of points which are feasible for $\mathbf{LP}(i,u)$ (i.e. the points that satisfy constraints (6)-(8)).

Proof: When all players other than i use u_{-i}, then player i is faced with a constrained Markov decision process (with a single controller). The proof of (i) and (ii) then follows from [5] Theorems 2.6. The first part of (iii) follows from standard properties of Linear Programs, whereas the second part follows from an application of the theory of sensitivity analysis of Linear Programs by Dantzig, Folkman and Shapiro [10] in [5] Theorem 3.6 to $\mathbf{LP}(i,u)$. Finally, (iv) follows from the definition of $g^i(z)$. ■

Define the point to set map

$$\Psi : \prod_{i=1}^N M_1(\mathcal{K}_i) \to 2^{\left\{\prod_{i=1}^N M_1(\mathcal{K}_i)\right\}}$$

by

$$\Psi(\mathbf{z}) = \prod_{i=1}^N \Gamma(i, g^i(z))$$

where $\mathbf{z} = (z_1, \ldots, z_N)$, each z_i is interpreted as a point in $M_1(\mathcal{K}_i)$ and $g(z) = (g^1(z_1), \ldots, g^N(z_N))$.
Proof of Theorem 2.1: By Kakutani's fixed point theorem, a fixed point $\mathbf{z} \in \Psi(\mathbf{z})$ exists. Proposition 3.1 (i) implies that for any such fixed point, the stationary multi-policy $g = \{g^i(z_i); i = 1,...,N\}$ is a constrained Nash equilibrium. ■

Remark 3.1 (i) The Linear Program formulation $\mathbf{LP}(i,u)$ is not only a tool for proving the existence of a constrained Nash equilibrium; in fact, due to Proposition 3.1 (ii), it can be shown that any stationary constrained Nash equilibrium w has the form $w = \{g^i(z_i); i = 1, ..., N\}$ for some $\mathbf{z}$ which is a fixed point of Ψ.
(ii) It follows from [5] Theorems 2.4 and 2.5 that if $\mathbf{z} = (z_1, ..., z_N)$ is a fixed point of Ψ, then any stationary multi-policy g in $\prod_{i=1}^{N} g^i(z_i)$ satisfies $C^{i,j}(\beta, g) = \mathcal{C}^{i,j}(z), i = 1, ..., N, j = 0, ..., B_i$. Conversely, if w is a constrained Nash equilibrium then

$$C^{i,j}(\beta, w) = \sum_{y \in \mathbf{X}} \sum_{a \in \mathbf{A}_i(y)} f_i(w; y, a) c_i^{j,w}(y, a)$$

(and $f(w)$ is a fixed point of Ψ).

References

[1] E. Altman, *Constrained Markov Decision Processes*, Chapman and Hall/CRC, 1999.

[2] E. Altman, K. Avrachenkov, R. Marquez and G. Miller, "Zero-sum constrained stochastic games with independent state processes", *Mathematical Methods in Operations Research*, Dec. 2005.

[3] E. Altman, K. Avratchenkov, G. Miller and B. Prabhu, "Uplink dynamic discrete power control in cellular networks", to appear in the proceedings of the 12-th International Symposium on dynamic games and applications, July 3-6, 2006, Sophia Antipolis, France.

[4] E. Altman and V. A. Gaitsgory, "Stability and Singular Perturbations in Constrained Markov Decision Problems", *IEEE Trans. Auto. Control*, **38**, No. 6, pp. 971-975, 1993.

[5] E. Altman and A. Shwartz, "Sensitivity of constrained Markov Decision Problems", *Annals of Operations Research*, **32**, pp. 1-22, 1991.

[6] E. Altman and A. Shwartz, "Markov decision problems and state-action frequencies", *SIAM J. Control and Optimization*, **29**, No. 4, pp. 786-809, 1991.

[7] E. Altman and A. Shwartz, "Constrained Markov Games: Nash Equilibria", Annals of the International Society of Dynamic Games, vol. 5, Birkhauser, V. Gaitsgory, J. Filar and K. Mizukami, editors, pp. 303-323, 2000.

[8] E. Altman and F. Spieksma, The Linear Program approach in Markov Decision Problems revisited, ZOR - Methods and Models in Operations Research, Vol. 42, Issue 2, pp. 169-188, 1995.

[9] R. Aumann and M. aschler, *Repeated Games with Incomplete Information*. M.I.T. Press, Cambridge, MA., 1995.

[10] Dantzig G. B., J. Folkman and N. Shapiro, "On the continuity of the minimum set of a continuous function", *J. Math. Anal. and Applications,* Vol. 17, pp. 519-548, 1967.

[11] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*, Springer, NY, 1996.

[12] E. Gómez-Ramírez, K. Najim and A.S. Poznyak, "Saddle-point calculation for constrained finite Markov chains". Journal of Economic Dynamics and Control, **27**, pp. 1833-1853, 2003.

[13] A. Hordijk and L. C. M. Kallenberg, "Constrained undiscounted stochastic dynamic programming", *Mathematics of Operations Research,* **9**, No. 2, May 1984.

[14] A. Hordijk and L. C. M. Kallenberg, "Linear programming and Markov games I", in *Game Theory and Mathematical Economics*, O. Moeschlin and D. Pallschke (eds.), North Holland, pp. 291–305, 1981.

[15] A. Hordijk and L. C. M. Kallenberg, "Linear programming and Markov games II", in *Game Theory and Mathematical Economics*, O. Moeschlin and D. Pallschke (eds.), North Holland, pp. 307–320, 1981.

[16] M. Huang, R. P. Malhamé and P. E. Caines, On a class of large-scale cost-coupled Markov games with applications to decentralized power control, IEEE CDC, Atlantis, Paradise Island, Bahama; Dec. 2004.

[17] M. Huang, R. P. Malhamé and P. E. Caines, Nash Strategies and adaptation for decentralized games involving weakly coupled agents, IEEE CDC, Dec. 2005.

[18] L. C. M. Kallenberg (1994), "Survey of linear programming for standard and nonstandard Markovian control problems, Part I: Theory", *ZOR – Methods and Models in Operations Research*, **40**, pp. 1-42.

[19] J. F. Mertens and A. Neyman, "Stochastic Games", *Int. Journal of Game Theory* Vol. 10, Issue 2, page 53-66, 1981.

[20] J. B. Rosen. Existence and uniqueness of equilibrium points for concave N-person games. *Econometrica*, 33:153–163, 1965.

[21] D. Rosenberg, E. Solan and N. Vieille, "Stochastic Games with Imperfect Monitoring", In Haurie A., Muto S., Petrosjan L.A., and Raghavan T.E.S., *Advances in Dynamic Games: Applications to Economics, Management Science, Engineering, and Environmental Management*, 2003.

[22] N. Shimkin, "Stochastic games with average cost constraints", *Annals of the International Society of Dynamic Games, Vol. 1: Advances in Dynamic Games and Applications*, Eds. T. Basar and A. Haurie, Birkhauser, 1994.

[23] Robert S. Simon, Stanislaw Spiez, Henryk Torunczyk, "Equilibrium existence and topology in some repeated games with incomplete information", *Trans. Amer. Math. Soc.*, 354 (2002), 5005-5026.

[24] O. J. Vrieze, "Linear programming and undiscounted stochastic games in which one player controls transitions", OR Spektrum 3, pp. 29–35, 1981.